8К

УДК 681.3

В.И. Кодачигов
Технологический институт Южного федерального университета, г. Таганрог, Россия
kodachigovvi@gmail.ru

Квазиканоническое кодирование графов Бержа

Предлагается метод получения квазиканонического кода графа. Такой код имеет существенно меньшую длину, чем известные канонический и универсальный коды. На основе предлагаемого метода легко строятся алгоритмы кодирования и декодирования с линейной оценкой сложности.

Мы рассмотрим только одну из самых простых разновидностей графов – графы Бержа, т.е. неориентированные графы без петель и кратных ребер, причем ребра графов имеют одинаковую длину, равную 1.

Цели кодирования:

1. Минимизировать длину кода (и, соответственно, время передачи графа по каналу связи).
2. Обеспечить конфиденциальность (скрытность) передачи.
3. Облегчить распознавание изоморфизма графов (РИГ) и т.д.

Итак, пусть задан граф G = (X, F_x), где X – множество вершин, F_x – отображение *X* в себя (т.е. множество ребер). Как известно, граф может быть представлен матрицей смежности *A*. Для графов рассматриваемого типа она симметрична и может иметь только 0 или 1 элементы. Поэтому вместо всей матрицы *A* можно рассматривать только ее верхнюю ($A_{\text{В}}$) или нижнюю ($A_{\text{Н}}$) треугольную матрицу. Выпишем элементы $A_{\text{В}}$ следующим образом в виде линейного списка. Сначала – первую строку, затем подряд – вторую, третью и т.д. Получим так называемый канонический код графа. Соответствующее ему число – это шифр графа. Спрашивается: можно ли нетривиально и взаимно-однозначно закодировать граф, в том числе и словом меньшей длины, чем канонический код. И если да, то какой ценой может быть достигнуто. Иначе говоря, какова (вычислительная) сложность, т.е. трудоемкость алгоритмов кодирования и декодирования графов?

Очевидно, что если коды двух графов, заданных соответственно матрицами *A* и *B*, совпадают, то определяемые ими графы G_A и G_B изоморфны. Однако если коды матриц *A* и *B* не совпадают, то это еще не значит, что G_A и G_B неизоморфны. Для окончательного установления факта изоморфизма надо сравнить код матрицы *A* с кодами всех матриц, изоморфных *X*.

Можно поступить иначе: привести *A* и *B* к некоторому стандартному, инвариантному виду и сравнить их коды.

А не являются ли инвариантами минимальные коды (шифры) матрицы? Быть может, если для двух графов G_A и G_B такие коды совпадают, то $G_A \sim G_B$ (здесь ~ – изоморфизм). Рассмотрим эту задачу более подробно. Итак, надо найти минимальные по длине коды (шифры), соответствующие произвольному графу.

Задача кодирования графа, в том числе и такого, что получаемый код меньше, чем канонический, впервые серьезно исследовалась в [1], [2]. Для этого предполагается использовать две процедуры – сжатие графов и универсальное кодирование. Для [1] важно определить, сжимаем граф в принципе или нет. Оказалось, что вообще говоря, это сделать можно не для любых графов, но есть классы сжимаемых графов.

В [2] предполагается и процедура универсального кодирования и декодирования графов. Сложность такого кодирования оценивается как $O(N^2)$, где N – количество вершин в графе, число же шагов декодирования не зависит от N. Для достаточно больших N длина кода сравнима с N^2.

Итак, нас интересуют следующие вопросы. Всегда ли универсальный код короче канонического? Не является ли универсальный код тем инвариантом, преобразования к которому позволило бы решить задачу распознавания изоморфизма?

Для ответа на эти вопросы предложенные в [2] методики были доработаны до алгоритмов, а сами алгоритмы были проверены на большом числе генерируемых случайным образом графов. Обработка результатов показала следующее:

1. Канонические (а также и универсальные) коды автоморфных графов в общем случае не совпадают.

2. Длина канонических кодов всех возможных автоморфных графов остается постоянной; его же числовое значение (шифр) тем больше, чем более качественно размещен кодируемый граф по критерию минимума суммарной длины ребер. Такому размещению соответствует и минимальный код. К примеру, обратимся к матрицам A и A', $A \sim A'$.Универсальный код минимальной матрицы A' имеет вид: K' = 01101111110110000111111, а код K графа с матрицы A таков: K = 1011111110110010100000001000001000010000.

3. Минимальной матрице соответствует максимальный шифр (из всех тождественных ей матриц).

Вывод 1. Универсальные коды, по-видимому, могут являться инвариантами решения задачи распознавания изоморфизма графов (РИГ), если они строятся для минимальных матриц. Для целей РИГ это ничем не лучше канонического кода.

Другое дело, если использовать универсальный код для сокращения длины кода. Такая процедура будет полезной в том случае, если преследуется цель минимизации времени передачи графа по последовательному каналу связи.

Можно ли сократить длину кода по сравнению с универсальным в последнем случае? Можно, если принять некоторое соглашение, известное лишь источнику и приемнику. Одним из таких соглашений может быть следующее: исходный граф считается размещенным по периметру окружности. Теперь вместо графа можно передать список, являющийся моделью так называемого стандартного графа. В [2] описано 5 моделей такого типа. Для примера одна из моделей стандартного графа первого типа, задаваемого списком U, такова:

U =
1 (5, 7)
2 (3, 6, 7, 8)
3 (2, 6)
4 (5, 6, 8)
5 (4, 7)
6 (8, 1, 2, 4)
7 (1, 2, 5)
8 (2, 3, 4, 6)

- это перестановка A_1=(16823754)

Вывод 2. Граф из n вершин можно взаимно-однозначно закодировать списком длины n.

Вывод 3. Для некоторых перестановок получаемое число-код длиннее самой перестановки. Представителем всех n! графов, автоморфных данному, является граф, определяемый минимальной матрицей.

Вывод 4. Минимальная матрица может считаться инвариантом графа *G*.

Итак, всякий граф может быть представлен в виде так называемого канонического кода. При отсутствии априорной информации о графе канонический код неизбыточен; если же мы имеем дело с графом из класса *X* так называемых наследственных графов [2], то граф может быть закодирован словом меньшей длины. Двоичный код графов из наследственного класса *X* имеет длину $\frac{1}{2}n^2 h(x)(l + E_n)$, где $E_n \to \infty$, при $n \to \infty$, $h(x)$ – предельное значение коэффициента сжатия, т.е. отношение длины кода, построенного с учетом принадлежности графа классу *X*, к длине канонического кода.

Любой непустой наследственный класс имеет более экономное кодирование, чем каноническое. В [1] описывается метод такого универсального кодирования.

Обратим внимание на то, что при универсальном кодировании фактически производится привязка графа к узлам двумерной решетки. Для этого кроме n надо задаться значениями *k*, *m*, *p*; *k* = [*n*/*p*]. Например, для *n* = 9: *k* = 3, *m* = 3, *p* = 3. Отметим, что кодирование (декодирование) графа можно определить с привязкой его вершин к более простому – круговому расположению.

В [2] предлагается несколько способов формирования такого так называемого стандартного расположения графов.

Обратимся к графу, которому соответствует список U(1).

Определим для пары смежных вершин *i* и *j* понятие расстояние

$$r_{ij} = |\min(+r_{ij}, -r_{ij})|,$$

где $+r_{ij}$ – число вершин, стоящих между смежными вершинами *i* и *j* при счете по часовой стрелке;

$-r_{ij}$ – число вершин, стоящих между смежными вершинами *i* и *j* при счете против часовой стрелки.

Заменим в списке *U* перечень смежных с текущей вершиной, *k* вершин расстоянием до *k* имеет список: 1(–2, –1). Построим по нему аналог канонического кода 2(0, 3, –2, –1), который назовем квазиканоническим кодом

U^* =
3(–2, –0)
4(0, 1, 3)
5(1, –0)
6(1, 2, 3, –1)
7(1, 2, –1)
8(1, 2, 3, –1)

–2, –1. 0, 3,–2, –1. –2, 0. 0, 1, 3. 1, –0.
1 2 3 4 5

1, 2, 3, –1. 1, 2, –1. 1, 2, 3, –1.
6 7 8

Здесь точки разделяют слова-коды (строки матрицы). Закодируем слово в 2-буквенном алфавите {0,1}. С учётом этого кодирования указанный выше квазиканонический код примет вид:

0000. 0010, 0011. 0101. 0110, 0111. 1000. 1001, 0011. 1001, 0100. 1001, 1010.
1 2 3 4 5 6 7 8

его длина равна 64.

Для сравнения: стандартный канонический код графа соответствующего *U*, 00111000.1111000.101000.10100.0001.001.10.1 его длина 44.

Как видно, получили код длиннее канонического.

Однако, такое кодирование избыточно, так как в списке U фактически одно и

то же ребро представлено и соответственно кодируется дважды как *ji* и *ij*.

–2, –1	0000
–0	0001
0	0010
3 –1	0011
–1	0100
–2 0	0101
0 1	0110
3	0111
1 –0	1000
12	1001
3 –1	1010
11	1011
21	1100

Преобразуем список U к безызбыточному виду

$$U_1 = \begin{matrix} 1(6,7) \\ 2(3,6,7,8) \\ 3(8) \\ 4(5,6,8) \\ 5(7) \\ 6(8) \end{matrix} \quad \rightarrow \quad U_1^* = \begin{matrix} 1(-2,-1) \\ 2(0,3,-2,-1) \\ 3(-2) \\ 4(0,1,-3) \\ 5(1) \\ 6(1) \end{matrix}$$

Закодируем слова в скобках в алфавите {0,1}:

–2 –1	000
0 3	001
–2	010
0 1	011
–3	100
1	101

Квазиканонический код для U^*

$$\underset{1}{\underline{-2,\ 1.}}\ \ \underset{2}{\underline{0,\ 3,\ -2,\ -1.}}\ \ \underset{3}{\underline{-2.}}\ \ \underset{4}{\underline{0,\ 1,\ -3.}}\ \ \underset{5}{\underline{1.}}\ \ \underset{6}{\underline{1}}$$

С учетом вышеуказанной таблицы имеем следующий код:

$$\underset{1}{\underline{000.}}\ \ \underset{2}{\underline{001,000.}}\ \ \underset{3}{\underline{010.}}\ \ \underset{4}{\underline{011,100.}}\ \ \underset{5}{\underline{101.}}\ \ \underset{6}{\underline{101.}}$$

Его длина 31. Напомним, что для канонического кода она равна 44. Можно заметить, что чем более связен и чем ближе к однородному исходный граф, тем экономнее кодирование.

Нетрудно построить алгоритм кодирования K. Его сложность, как видно из его построения, равна $O(Cn)$. Поскольку все операции в K взаимно-однозначные, то легко построить и алгоритм декодирования; его сложность также имеет оценку $O(Cn)$.

Литература

1. Алексеев В.Н. О сжимаемых графах / В.Н. Алексеев // Проблемы кибернетики. – 1979. – № 36.
2. Алексеев В.П. Наследственные классы и кодирование графов / В.П. Алексеев // Проблемы кибернетики. – 1982. – № 39.
3. Кодачигов В.И. Стандартные графы и их модели / В.И. Кодачигов // Известия ТРТУ. – Таганрог, 2005.

В.І. Кодачигов

Квазіканонічне кодування графів Бержа

Пропонується метод отримання квазіканонічного коду графа. Такий код має істотно меншу довжину, чим відомі канонічний і універсальний коди. На основі пропонованого методу легко будуються алгоритми кодування і декодування з лінійною оцінкою складності.

V.I. Kodachigov

Quasicanonical Coding of Berge's Graphs

The method in order to arrive the quasicanonical code of graph is present. This code is significantly shorter then the known canonical and universal codes. On the basis of the proposed method it is easy to construct algorithms for encoding and decoding with linear estimation of complexity.

Статья поступила в редакцию 01.06.2010.